# Automatic Tagging Of Turns in the London-Lund Corpus with Respect to Type of Turn

**Benny Brodda**
**Stockholm**

## 0. Abstract.

In this paper a fully automatic tagging system for the dialogue texts in the London-Lund corpus, LLC, will be presented. The units that receive tags are "turns"; a collection of (not necessarily connected) tone units – the basic record in the corpus – that one speaker produces while being either the "floor holder" or the "listener"; the quoted concepts are defined below. The tags constitute a classification of each turn according to "type of turn". A little sample of tagged text appears in Appendix 1, and is commented on in the text. The texts to be tagged will in the end comprise all the texts in the three subcorpora of LLC appearing in Svartvik & Quirk, "A Corpus of English Conversation", (=CEC); so far, about half of these texts have been tagged, now with the programs working properly, the rest will hopefully be tagged before the end of this year.

## 1. Introduction

An outline of the classification scheme underlying the present tagging system was presented in Brodda, 1988, and is essentially the same classification system used in this report. In the present project, however, the classification is made explicit through the tags, simplifying the verification problem considerably.

The tagged texts will provide a basis for a statistical investigation of the corpus; one important question that will be adressed is whether or not speakers tend to differ in the factors these tags reflect when the speakers sex, social rank, or other properties that CEC provides about the participants of each dialogue text are taken into account. Britt Erman, Stockholm, will present a linguistic investigation of such factors in the same corpus. By the end of this year, we hope to have the statstical evaluation completed.

The underlying factors the tags reflect are probably to some degree semantic, sociolinguistic and context dependent, but primarily they show considerable individual variability related to the participants speaking habits, to their mental or physical mood at the recording occasion, the topic that happens to be discussed and so on. This means that one will have to take a considerable number of texts into account in order to filter out such individual variations, while hopefully retaining some significant



Proceedings of NODALIDA 1993, pages 51-63

residual. A sample of the of basic frequencies that will go into the statistical machinery is presented in appendix 3.

The corpus itself, as well as the programs involved in the project are adapted to an ordinary (IBM compatible) PC-environment. Once the corpus is modified as described in section 2 below, the final tagging of each text will take about one minute on a 286-PC (16 Mz) and about the same time for frequency counts.

## 2. Corpus preparation.

In order to get the tagging and statistical programs working properly, a substantial simplification and standardisation of the corpus itself has been carried out. Elsewhere I will present a critical and thorough analysis (Brodda, 1994) of the LLC corpus and its technical design as it has usually been distributed to research groups around the world. That report will also contain a full account of the general purpose modifications made for the present project. It is quite clear that one can simplify the texts considerably, without loss of any information whatsoever, and at the same time better suit them for automatic analysis by computer. The fact that the revised corpus requires less than half the disk space of the original text (still in pure ASCII) is probably good news as well, especially when working on a PC where diskspace is not always an unlimited resource.

### 2.1 The basic modification of the corpus.

The basic record of the corpus is still the T(one) U(nit), but it has now a more BROWN-corpus style structure:

```
Text-id TU-id Speaker-id t....e....x....t#
```

where the line headers here are of length 14 ("line header" = a fixed length, initial portion of each line not containing the text itself). The tone unit delimiters, "#", are moved to the actual ends of complete TUs, meaning they become formal end markers of complete basic units. A "~" is inserted as a corresponding end marker of each incomplete TU.

The texts are then sorted in ascending order with respect to the TU-id(entifier)s in such a way that the lines constituting one TU appear in the sorted text in the same relative order as in the original text; a "+" is prefixed the Speaker-id to indicate that the TU continues on the next line. In this sorting it is chunks of simultaneous speech that are shuffled around, but simultaneous speech represents nonlinear events, anyhow, so the sorted text is absolutely equivalent to the original text; cf. p. 6 in the foreword of CEC. The sorting makes the text considerably less fragmented.

For the specific investigation presented here we did not need the prosodic markers, so we simply rinsed the text of these (which saves another 8% disk space). If this rinsing is done carefully, every word form can be rendered a "stable" spelling, which simplifies any type of parsing of the corpus (a simple parse is employed in the present project). Later we will try to see if the classification can be refined, when prosodic markers are taken into account, or if the tags correlate with these prosodic markers in one way or other.

The mentioned modifications of the corpus are all completely done by computer. We have also made a few (semi)manual modifications in order to standardise the texts further still; this standardisation is "general purpose" (not tied to this specific project), and should simplify any type of automatic analysis of the corpus; cf. Brodda, 1994.

## 3. A brief description of the turn classifying algorithm.

### 3.1 Turns and Formal Turns.

Let us start with a little exposé of things familiar to everyone and included in order to pinpoint a few phenomena that my programs identify.

Usually one "turn" in a dialogue is conceived as a stretch of speech that one participant utters in a connected sequence of words, phrases, tone units, or whatever elements speech is assumed to be made up of. In well disciplined dialogues each participant is allowed to deliver his/her turns uninterrupted; when a participant finishes his/her turn, another "takes the floor", delivers his/her turn, and the dialogue procedes in an orderly fashion. These kinds of turns I call "regular turns", and the switching between them I call "regular turn taking".

In more informal dialogues people are not that well behaved. Participants laugh, start talking when someone else already has the floor, and so on. Sometimes these are simply side comments to what the floor holder is saying, at other times the new speaker brutally takes over the floor (we have a "takeover" situation), perhaps accompanied by an increase in voice volume. Sometimes the takeover fails (perhaps the floor holder raises his/her voice still more and manages to maintain the floor); such a situation I call an "attempt".

Even when participants are disciplined and await their turns in an orderly fashion, they are not always silent, at least not in less formal situations. They deliver typical feedback signals: "yeah", "of course", "yes", "certainly", and they laugh etc., sometimes while the floor holder is actually speaking (in which case I call such signals "back channels"), or when the floor holder briefly pauses (in which case I call them "feedbacks"). Such feedback signals do not break the floor holders turn, and indeed are not meant to. In most everyday dialogue situations, such interaction is, in fact, quite necessary – in telephone dialogues it is mandatory – and has a purely supportive function. In more formal situations, such as seminars and the like, head knoddings, smiles and so on, have this same function.

The auxiliary "formal turn" concept below, is a first approximation of a more final turn concept; I will return to this later. The formal turns will be the object for the tagging algorithm.

The **formal turn** concept:
A formal turn (**FT**) is the collection of all TUs in a maximal, unbroken sequence of TUs assigned to one and the same speaker; "maximal" in the sense that the FT cannot be further extended and still be an FT (i.e. the FT in question is surrounded on both sides by either some other speaker's FTs or text end-markers). The term "speaker" here means the string constituting the Speaker-id field; thus speakers "A" and "AB" ("A" or "B") are distinct. Cf. section 3.3 about speaker ",".

In the text samples below, formal turns are identified by the FT end marker, "|".

## 3.2 The tagging scheme

Table 1 below, summarises the tagging scheme. Each formal turn receives a turn type tag, viz. any of the characters in the set TC of tagging characters:

```
TC = {r, c, t, a, l, m, f, b, u, ","};
```

The classification algorithm runs in two passes. In the first pass the FTs are classified "context free". In this pass all types of turns except the "c"-turns are provisionally recognised. In the second pass, which employs a kind of context sensitive rules, some of the tags from the first pass are changed in one way or another; primarily the "c"-tags that are now introduced.

Table 1. The Turn Type Classification Scheme

```
S represents the speaker,"|" is the formal turn endmarker.

regular turn:        S    r    bla bla *bla* bla|
continuation:        S    c    bla bla bla
takeover:            S    t    *bla bla* bla bla|
attempt:             S    a    *bla bla*|
feedback signal:     S    f    yes|
backchannel:         S    b    *yes*|
laugh:               S    l    (laughs)|
back ch. laugh:      S    m    *(laughs)*|
human noise:         S    u    (cough)| or *(cough)*|
external noise:      ,    ,    (bang)| or *(bang)*|

"attempt" is short for "floor stealing attempt";
"take over" is short for "brute force floor take over";
"*" denotes a "break character"; cf. below.
```

### 3.3 The Context Free pass.

The first pass recognises explicitly the l, m, f, b, u and "," turns according to what the FTs contain as indicated in Table 1. Thus an FT receives the tag "f" (= feedback) if it contains a mere "yes" or any other more or less synonymous word according to a little lexicon containing some 15 odd elements: ("yeah, mm, quite,.."), and if it is not enclosed in break characters (cf. immediately below); even FTs containing a combination or repetition of these elements receive this tag:

```
"A f oh yes yes yes|" (A is the speaker)
```

The same FT will receive the tag "b" if it is enclosed in a pair of "break characters", any of the characters or character combinations "*", "+", "**" or "++". Thus the following is a typical b-turn:

```
"A b *oh yes yes yes*|"
```

Break characters come in quadruples. A pair, like the one above, indicates that A's utterance is produced while someone else is talking. Immediately above or below this b-turn there should occur another FT containing a stretch of speech enclosed in the same pair of break characters, indicating that the matching stretches of speech occur simultaneously. (Cf., e.g., TUs 38 and 39 in the text sample in Appendix 1.)

The l, m, u and "," turns are likewise recognised through lexical lookups; thus, turns receive the "l" or "m" tags if the turns solely contain strings like "(laughs)", "(giggles)" or a few variants of these. The ","-turns are those FTs that appear in the original corpus without a speaker-id(entifier), typically indicating an external noise of some kind, such as "(phone rings)", "(car noise)", etc. (In the modified corpus the comma is also used as speaker-id for such TUs, meaning that each TU in the corpus formally has an owner.) The FTs may also contain various combinations of the elements mentioned above, and then they receive a tag according to a kind of heuristic rules. Thus, an FT of the type "A (laugh) yeahl" gets the "l"-tag, the FT "A yeah (laugh)l" the "f"-tag.

Every FT not explicitly recognised in this first pass is considered to contain "real" – more substantial – speech ("bla bla" in Table. 1). Thus, real speech is negatively defined. FTs containing real speech receive any of the "a", "t" or "r" tags depending on whether the FT contains simultaneous speech in a dominant way or not. If the FT is completely enclosed in break characters, it receives the "a" tag, if it only has an initial part enclosed in such characters, it receives the "t"-tag, otherwise it receives the "r"-tag.

### 3.4 The context sensitive pass

In the second pass the following explicit assumption is built into the program:

The **floor holder** concept:
At any moment in time (at any place in the text actually, from the program's point of view) there is always one dialogue participant that is established as the **floor holder, FH**. There are exactly two ways the FH may shift, viz. through what I call **significant** turn taking events.

The program assumes an "unspecified" speaker – distinct from all actual participants – as holding the floor when a text begins.

#### 3.4.1 Significant turn taking events.

One way the FH may shift is through **regular turn taking**: A speaker, other than the established FH, enters and delivers an FT that has been classified as an "r"-turn in the first pass. The owner of this new "r"-turn then becomes the new FH and the FT retains its "r"-tag.

Another way the floor holder may shift is through a **takeover**. This situation – a typical example of which appears in the text sample in Appendix 1 at TU 30 – occurs when the established FH's latest FT ends in a stretch of simultaneous speech that overlaps with simultaneous speech in the beginning of a new speakers FT, which has been given the "t"-tag in the first pass; these stretches of simultaneous speech must also contain "real speech". The actual floor holder shift takes place – the program assumes – precisely at the point where the inital stretch of simultaneous speech ends in the new speakers FT. In a takeover situation, this FT retains the "t"-tag from the first pass.

Both "t"-tags and "a"-tags may sometimes be changed into "r"-tags. This happens when i. the owner of the corresponding FT differs from the established FH, and ii. the prominent stretch of simultaneous speech in this FT matches a stretch of speech that is of a "weaker" category than the present in a turn type strength hierarchy, TSH, implicitly reflected in Table 1 but more formally defined as:

```
TSH: r > t > a > f > b > l > m > u > ",";
```

where the symbol ">" (here) stands for the two-place predicate "is stronger than". For present purposes, only the order between "t", "a" and the weaker ones is of interest. Thus, if the prominent stretch of simultaneous speech in a "t"-tagged FT matches that of an "a"-tagged or weaker, then the "t"-tag is turned into an "r"-tag; an "a"-tag is similarily turned into an "r"-tag, if it matches an "f"-turn or weaker.

The full hierarchy is needed for describing certain details of the statistic calculations.

### 3.4.2 Continuations

A typical episode in a dialogue starts with a sequence of "r"-turns, i.e. the floor holder shifts regularly from one speaker to the other. If any of the turns of the a, f, l, b, m, u or "," types are encountered, the floor holder normally does not shift, and let us assume now that he does not.

After such an interlude, two things may happen. Either the floor holder reappears in the FT immediatly following such an interlude, or a third participant appears (remember, a shift in FT always implies a shift in speakers). In the first case this new FT receives the "c"-tag regardless of what tag it received in the first pass, and it is assumed to be a continuation of the same speaker's former turn. If another speaker appears immediately after the interlude, then this new FT is treated as any other new FT as described above, i.e. the new speaker may become the new floor holder or the corresponding FT is just another interlude.

## 3.5 Turns

The second pass is considerably more complex than I have indicated here. Among other things, certain FTs are broken up into sub-FTs as indicated through a "\" in the text samples below. Many more details could be commented upon, but I think we are ready to define a final "turn" concept.

Major turns:
A **major turn** is a collection of FTs assigned to one and the same speaker, beginning with a significant turn taking event and interrupted only by such FTs that do not imply a shift in floor holder. Thus, a major turn always begins with either an "r" or a "t"-tag, i.e. when the speaker enters the floor, and zero or more "c"-tagged FTs that are continuations of the same turn. The whole turn is called a "regular" turn or a "takeover" depending on the tag on the initial FT.

Minor turns:
A **minor turn** is a formal turn that has any of the tags in the subset {a, f, b, l, m, u, ","} of TC after the second pass.

At any moment in time, the established FH is the speaker (at that moment) and the other participants are the (temporary) listeners.

## 4. Illustrations

What is described in section 3 above is of course a computer model intended to capture certain aspects of turn taking in the LLC-texts (or in any informal dialogue), and as any such model it captures reality more or less good. The evaluation so far, indicates very good corrolation between how the computer classifies turns in the LLC corpus and how students at the English department at Stockholm University do it. There is not enough space here to present larger samples of tagged text, but the samples given in Appendixes 1 and 2 would at least give an indication of what the tagging looks like.

The mentioned text sample illustrates a typical episode in a longer dialogue. After B's initial "r"-turn, speaker A starts an "r"-turn at TU 26 but encounters a prototypic takeover by B (the shift from TU 29 to 30). B manages then to keep the floor all the way down to TU 48. Thus, B's "t"-turn consists of the FTs (identifying each FT through its initial TU-id) 30, 34, 38, 40, 43 and 47.

Note, this takeover is also a semantic takeover. When the episode begins they are involved in a discussion about A's years as a young student, a topic that A continues to evolve in FT 26. B, however, breaks in and starts talking about her own years as a young student. (Both speakers are female).

In Appendix 2 a few special cases are given. Ill. 2.1 illustrates an interesting error of principle. FT/TU 1042, which consists of a single "yes#", is – precisely according to the algorithm – given the "f"-tag, i.e. classified as a feedback signal. If one scrutinises the context more closely, it appears, however, that this "yes" is an affirmative answer to a straightforward yes/no-question. According to any linguistic criteria it must, of course, be considered as a substantial turn; it adds semantic material to the dialogue and should be given the "r"-tag.

The text L1-5 contains about 115 FTs consisting of a single "yes", "yeah" or "yea" (enclosed in break characters or not). As far as one can deduce from the text, every one of them except the mentioned TU 1042 are feed back signals and not substantial turns (and consequently correctly tagged by the program). Text L1-5 is quite representative for the informal dialogues in the LLC-corpus, and the investigation so far seems to indicate that only about 1% of all single "yes"es produced in such dialogues represent substantial turns.

What about "no"? FTs consisting of a single "no" are, of course, considerably fewer than those consisting of a single "yes". Contrary to what one may think, though, also "no"-turns tend to be feedback signals more often than substantial turns, and they are regularly so when produced in some negative context, in which case they indicate that the no-sayer agrees with what is just said; strictly speaking "no" then means "yes" (cf. FT 1187 – Ill. 2.2 – where speaker A says both "no" and "yes"). Such a "no"-turn I call an "affirmative no". The tagging program assigns a "b"- or "f"-tag to a "no"-turn, if the preceding FT simply contains the word "not" or the word end "-n't" regardless of context, and so far this simple surface criterion has never gone wrong.

Appendix 3 contains the type of frequency tables that will underlie the statistical evaluation. The tables describe speaker A's "event history" during the whole dialogue l1-5. The first two tables show what speaker A does, how many turns of different types she produces, and the total no. of words produced during each turn type.

The last two tables describe the kinds of "attacks" speaker A encounters while holding the floor, i.e. the types of simultaneous speech other participants produce while A is the FH, and the number of words A produces during these attacks. (Some of these "attacks" are certainly not real attacks, since the feedback types are "supportive" rather than "hostile").

In the same way we obtain corresponding figures for every single speaker in any of the texts in the corpus, which figures then are inserted into a database, together with information about the speaker's sex, the number of participants in the corresponding dialogue and the other participants' sex. This database will provide the basis for a statistical investigation of the corpus.

## References

Brodda, B. 1988. *Tracing Turns in the London-Lund Corpus with BetaText.* JLLC, Vol. 3, No.2.

Brodda, B. 1994. *Simplifications and Modifications of the LLC-corpus in Preparation for Automatic Analysis Internal report,* Department of Linguistics, Stocholm University, S-10691, Stockholm, Sweden.

Erman, B. 1994. *Computer Analysis of Female and Male Conversational Strategies in same-sex and mixed-sex interaction in LLC,* Paper sent to the ALLC-ACH94 conference.

Svartvik, J and R. Quirk. 1980. *A Corpus of English Conversation,* London Studies in English, Lund.

## Appendices: Illustrations

## Appendix I: Sample tagged text (from L1-5T.TXT):

```
==========================================================
1 5 25+B    r    I don't suppose you need Old English and
1 5 25 B         Anglo-Saxon#|
1 5 26 A    r    well no .#
1 5 27 A         but [@m] you know#
1 5 28 A         *I don't#
1 5 29 A         have any language*#|
1 5 30+B    t    *[@m] well I <hadn't>* done any English at
1 5 30 B         **all**#
1 5 31 B         you know#
1 5 32 B         since O-level .#|
1 5 33 A    f    **<1 syll>** yea .#|
1 5 34 B    c    and I went to some second year {seminars}#
1 5 35+B         where there are only about half a dozen
1 5 35 B         people#
1 5 36+B         *and* they discussed what <a>
1 5 36 B         word was#|
1 5 37 A    b    *[m]*#|
1 5 38 B    c    **and -** what's a sentence#|
1 5 39 A    b    **[m]**#|
1 5 40 B    c    that's *even* more difficult .#|
1 5 41 A    b    *yeah*#\
1 5 42 A    f    yeah -#|
1 5 43 B    c    and so on .#
1 5 44+B         and then I also went to some postgraduate
1 5 44 B         ones#
1 5 45 B         which were more interesting -#|
1 5 46 A    f    yea#|
1 5 47 B    c    which he had for [dhi] - diploma -#
1 5 48 B         the main people#|
1 5 49 A    r    on -#|
1 5 50+B    r    and I suppose they're doing the same
                 ones this
1 5 50 B         year#
1 5 51+B         and then you'd have a whole evening
                 {battling
1 5 51 B         on} - - -#|
1 5 52-?    r    <4 to 5 sylls> - --|
1 5 53 B    r    no#
1 5 54 B         sessions .#
1 5 55 B         several sessions#
1 5 56 B         on *nominal* groups or something#
1 5 57+B         <then> you can
1 5 57 B         pick up all the jargon#|
1 5 58 A    b    *[m]*#|
1 5 59-B    c    and~|
1 5 60 A    f    yea - -#|
1 5 61+B    c    and then sort of get the hang of
                 what they're
1 5 61 B         talking about -#
1 5 62+B         I should ask him {if there are any
                 seminars you
1 5 62 B         ought to go to}#|
1 5 63 A    f    yea -#|
```

## Appendix 2:

==================================================

### ILL. 2.1: An "f" that is a real turn:

```
---------------
1 5 1037 C    c   I mean I've worked in universities#
1 5 1038 C        for nearly ten years now#|
1 5 1039 A    f   yeah .#|
1 5 1040 C    c   *and*#|
1 5 1041 A    t   *are* you going to America#|
1 5 1042 C    f   yes#|                               <------
1 5 1043 A    c   [m]#
1 5 1044 A        I [z] . tried to go to America#
1 5 1045 A        earlier this year#
1 5 1046 A        *and* then decided <syll syll>#|
1 5 1047 C    b   *[mhm]*#|
```

### ILL. 2.2: Examples of "affirmative no"

```
---------------
1 5 779+C     c   [@] - - but [@] they're just sort
                  of pursuing
1 5 779 C         their own research#|
1 5 780 A     f   yea#|
1 5 781 C     c   they're probably teaching elsewhere#|
1 5 782 A     f   yea#|
1 5 783 C     c   . and [@] they don't seem to
                  bother anybody#|
1 5 784 A     f   no#|                                <------
1 5 785 C     c   they seem to know their way around#|
1 5 786 A     r   so it does seem#
1 5 787 A         a fairly self-contained *unit on
                  its own*#|

-----------
1 5 1175 A    t   *I'm* also#
1 5 1176 A   [t]  reasonably anxious#
1 5 1177 A        to bump into people#
1 5 1178+A        but perhaps one just . sort of - holds on
1 5 1178 A        that -#|
1 5 1179 D    r   well yes#
1 5 1180+D        that's - - - it's not so easy as .
                  you think
1 5 1180 D        *really*#|
1 5 1181 A    b   *no*#|                                <------
1 5 1182 D    c   because - being over here#
1 5 1183 D        we tend to be a bit isolated#|
1 5 1184 A    f   yeah#
1 5 1185 A        [m] - -#|
1 5 1186+D    c   [m] specially as we don't go to . to
                  coffee
1 5 1186 D        over in [dhi] . *the main building
                  you see*#|
1 5 1187 A    b   *no .#                                <------
1 5 1188 A        yes*#\
1 5 1189-A    r   that's what~|
```

## Appendix 3: Sample frequency counts: Text: L1-5T.TXT

```
=========================================================

Speaker A
Number of turns:
AFT:   309    total no. of formal turns prod. by A
ArT:    68    no. of turns produced by A
AtT:    16    no. of t-turns produced by A
AaT:     8    no. of a-turns produced by A
AfT:    74    no. of f-turns produced by A
AlT:     3    no. of l-turns produced by A
AbT:    92    no. of b-turns produced by A
AmT:     4    no. of m-turns produced by A
AuT:     0    no. of u-turns produced by A
&&

Number of words produced by A
ATW:  1786    total no. wrds produced by A
ArW:  1384    no. of words prod. by A as floor holder
AtW:   264    no. of words prod. by A during t-turns
AaW:    19    no. of words prod. by A during a-turns
AfW:    53    no. of words prod. by A during f-turns
AlW:     0    no. of words prod. by A during l-turns
AbW:    65    no. of words prod. by A during b-turns
AuW:     0    no. of words prod. by A during u-turns
&&


Number of "attacks"
ATt:    55    total no. attacks on A when being FH
ArT:     3    no. of attacks of type r
Att:    14    no. of attacks of type t
Aat:    14    no. of attacks of type a
Aft:     2    no. of attacks of type f
Alt:     0    no. of attacks of type l
Abt:    17    no. of attacks of type b
Amt:     4    no. of attacks of type m
Aut:     0    no. of attacks of type u


Number of words prod. by A during attacks
ATw:   114    total no. wrds produced by A during attacks
Arw:     2    no. of words prod. by A during r-turn attacks
Atw:    30    no. of words prod. by A during t-turn attacks
Aaw:    58    no. of words prod. by A during a-turn attacks
Afw:     2    no. of words prod. by A during f-turn attacks
Alw:     0    no. of words prod. by A during l-turn attacks
AbW.    18    no. of words prod. by A during b-turn attacks
Amw:     3    no. of words prod. by A during m-turn attacks
Auw:     0    no. of words prod. by A during u-turn attacks
```